

WALDEMAR ROJAS, Ed.D.
Superintendent of Schools

San Francisco Unified School District
135 Van Ness Avenue, San Francisco, CA 94102



Thursday, April 15, 1999

INSTITUTE OF GOVERNMENTAL
STUDIES LIBRARY

APR 27 1999

To: SFUSD Stakeholders

Subject: SFUSD Testing Policy

UNIVERSITY OF CALIFORNIA

Recently, our testing and student achievement reporting practices have been unfairly maligned by the local press. We **want** to assure you that SFUSD's testing policies are completely aligned with the American Educational Research Association, The American Psychological Association, and the National Council on Measurement in Education standards.¹ Newspaper reporters are *not* educational psychometricians; therefore, we would like to outline some of the fallacies and omissions put forth by the *San Francisco Examiner*² on the way we measure student achievement in our district.

The San Francisco Examiner states: "...an Examiner analysis of student achievement data shows that behind the vaunted success is an eye-opening district practice: While student enrollment remains flat, test results from fewer students are being reported each year, raising questions about whether the gains are real or manipulated."³

THE FACTS: We test all students who are eligible to be tested and are present on the day of the test in grades 1-11 on the CTBS and in grades 2-11 on the SAT-9 test.

The Examiner says that we exclude from testing "a large segment of the lowest-performing kids..."⁴

THE FACTS: There are two groups of students we do not test: Limited English Proficient (LEP) students with less than 30 months of English instruction, and Special Education (SPED) students who are exempt by law from testing. To test these groups of students would render results that are not valid or reliable. Testing experts agree, "If English-language learners do not understand test questions due to unfamiliarity with the language of the test, interpretations of their scores as showing their actual achievement will not be valid."⁵ The California Department of Education adopted these guidelines in 1993. They are consistent with the National Assessment of Educational Progress standards on which children should be tested. When the State changed its guidelines in 1998, the district defended LEP students' civil rights and was victorious in every Court, including the California Supreme Court. While the courts and the State Legislature have upheld the practice of exempting these LEP students, and the Examiner admits that the District has done, "nothing illegal or unethical," the article fails to state that testing these LEP students would be unproductive and unethical.

The Examiner states: "Although there is nothing illegal or unethical about excusing children who lack English skills from taking English-only exams, it puts San Francisco at odds with other school districts nationwide..."⁶

THE FACTS: The SFUSD has always tested more than 90% of its eligible students. The rationale for testing our students is not to "compete" with other districts, but rather to tell us about students' academic growth and needs. Rather than subjecting our LEP students to high-stakes testing that does not produce useful data, we use other educationally-sound

1. William J. Russell, Ph.D., Executive Officer, AERA; Raymond D. Fowler, Ph. D., Chief Executive Officer, APA; and Edward H. Haertel, Ph.D., President, NCME; personal communication, May 15, 1998.
2. Guthrie, Julian (April 12, 1999). S.F. test scores questioned: Improvement may be smoke and mirrors. *San Francisco Examiner*. pp. A-1, A-10.
3. (see footnote 2)
4. (see footnote 2)
5. (see footnote 1)
6. (see footnote 2)

THE UNIVERSITY OF CHICAGO
LIBRARY

APR 27 1978

1100 S. MICHIGAN AVE.
CHICAGO, ILL. 60607

approaches for measuring their growth, like the SABE test in Spanish. In addition some LEP students with less than 30 months in the district *are* tested in English, if they are recommended for testing by their teachers. Most importantly, the district is now implementing another assessment system for second language learners which measures their achievement in two languages, at multiple proficiency levels, and across the four areas of language learning: listening, speaking, reading and writing.

The Examiner states " . . .many students' scores no longer are reported."

THE FACTS: We publish *all* scores every August in the *SFUSD Academic Achievement Report*. This is available to the public and the media through our Research, Planning and Evaluation Office (telephone: 415-241-6454). Our testing data is also available at our **website**: www.sfusd.edu.

The Examiner states: " . . .the district primarily reports the scores of those students who have taken the test at the same school for two consecutive years. Those scores are called 'matched scores'...the district defends its practices."

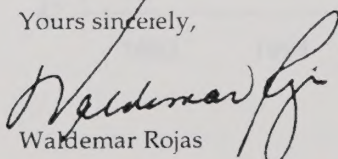
THE FACTS: SFUSD *does report all* scores, and these are readily available to the general public. However, we use *matched* scores because they are better diagnostic tools for our educators; they help us to pinpoint instructional needs. Matched scores are the scores of students who have been in the district for two or more consecutive years. This ensures that we are not comparing children who have been educated elsewhere, and who have just entered the district, to those who have been educated through our programs for at least two years. We *do* examine all scores to see how we compare nationally; however, the matched scores are more useful for us in implementing policy changes that produce real student achievement. *Anyway you look at our test scores, matched or unmatched, the results show the same positive trend: Our students are improving.* In fact, if we truly wanted to "inflate" our scores, we would test **the** LEP students with less than 30 months of English because they show the greatest gains from year to year!

The Examiner states: "Between 1987 and 1998, Chinese American enrollment increased by nearly 5 percent, while black enrollment declined by nearly 4 percent...On the district exam...Asian Americans score significantly higher than students of other ethnic and racial groups. African Americans score the lowest."

THE FACTS: First of all, either our students are improving or they are not. The press can't have it both ways. If Asian population growth in San Francisco is driving our improvement, then Asian growth all over the state similarly affects scores in other districts. However, not all Asian students are overachievers, nor are all Latino and African American students underachievers. We question the *Examiner's* use of ethnic stereotypes. In fact, we have many high-achieving students of all races, and most of our underachievers **have** a single characteristic in common: poverty. Nonetheless, we in the SFUSD believe that all children can achieve with proper instruction and a good learning environment.

We regret that the work of our teachers **and** administrators toward raising student achievement has been shown such a total lack of respect. We are proud of the gains we **have** made, and we thank all our stakeholders for their continuing work. Perhaps the reporter simply did not understand the testing data; nonetheless, this is no excuse for biased reporting.

Yours sincerely,



Waldemar Rojas

The first part of the report is a summary of the work done during the last year. It is a very brief summary, but it gives a good idea of what has been achieved. The second part of the report is a more detailed account of the work done during the last year. It is a very detailed account, but it is also very interesting. The third part of the report is a summary of the work done during the last year. It is a very brief summary, but it gives a good idea of what has been achieved.

The following table shows the results of the work done during the last year.

The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting.

The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting.

The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting.

The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting.

The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting.

The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting. The table shows the results of the work done during the last year. It is a very detailed account, but it is also very interesting.

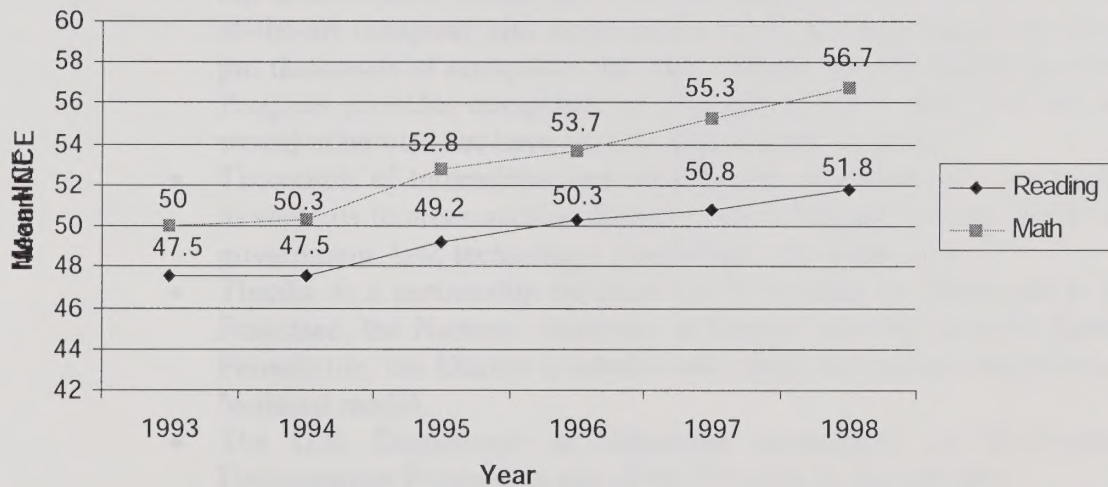
Signature
Date: _____

District Test Scores Are Rising, Are San Francisco Schools Really Getting Better?

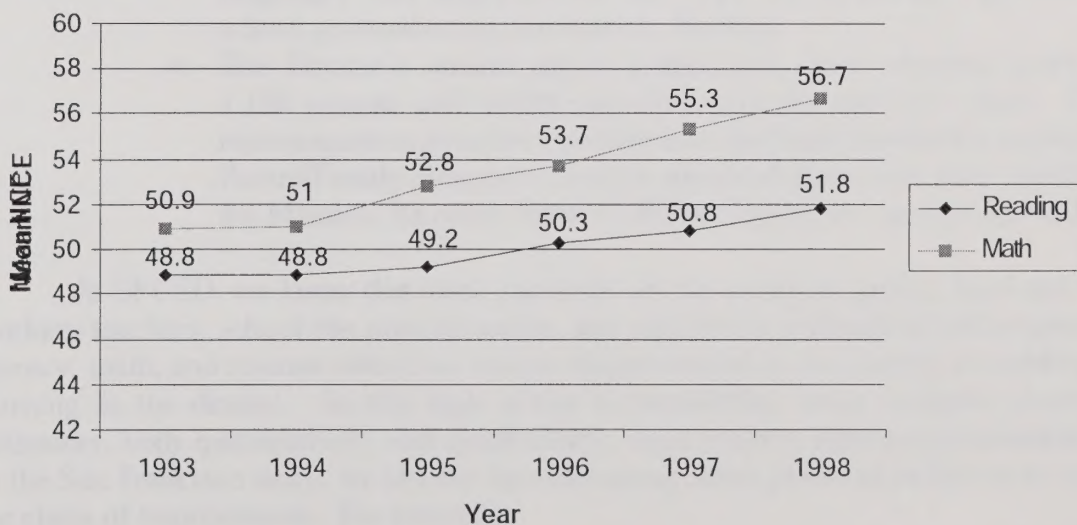
In the past six years, standardized test scores and other measures have improved significantly in San Francisco schools:

- Student reading and math scores on the Comprehensive Tests of Basic Skills (CTBS) reflect an unprecedented increase for the sixth year in a row, and students' matched scores exceeded the 50th percentile for the third consecutive year on both sub-tests of the CTBS (See Chart 1 and 1A).

Chart 1: District CTBS Scores 1992-93 to 1997-98 Grades 2-11



**Chart 1A. District CTBS Scores 1992-93 to 1997-98 Grades 2-11
(1992-93 and 1993-94 adjusted for LEP<30 implemented in 1995)**



Are San Francisco Schools Really Getting Better? Student Test Scores Are Rising

In 1990, the first, standardized test scores and other measures have improved significantly in San Francisco schools.

• Student test scores rose in 1990 for the Comprehensive Tests of Basic Skills (CTBS) in the areas of reading, mathematics, and science. The scores rose in all three areas, with the largest gains in reading and mathematics. The scores for the CTBS in 1990 were 51.7 in reading, 50.7 in mathematics, and 50.7 in science. The scores for the CTBS in 1989 were 50.7 in reading, 50.7 in mathematics, and 50.7 in science.

(A)

Chart 1: Student CTBS Scores 1989-90 to 1990-91 (Grade 5-11)



Chart 2: Student CTBS Scores (Grade 5-11) 1989-90 to 1990-91 (Grade 5-11)



- Over this same time period, the number of students performing in the top quartile has steadily increased (19% in 1993 to 24% in 1998), while the number of students in the bottom quartile has consistently decreased (34% in 1993 to 24% in 1998).
- From 1993 to 1998, in part as a result of significant professional development in literacy, math and science, and reducing class size from 30 to below 20, the first grade test scores in reading rose from 34.6 to 48.3 NCEs, second grade went up from 38.4 to 49.7 NCEs, and third grade (maximum class size of 25) scores went up from 45.7 to 49.1 NCEs.
- Our schools consistently win recognition at the state and national level, including the California Distinguished Schools Program, Redbook's Top Schools in the Nation and the U.S. Department of Education's Blue Ribbon School Award.
- All schools have e-mail and Internet capability and many have state-of-the-art computer and multi-media labs. Corporate partners helped put thousands of computers into the schools, and the Home Computer Program provides computers to students, families, and teachers who would otherwise not have access to technology at home.
- Thousands of internships and employment opportunities are available to students in areas such as finance, biotechnology, travel, health, city government, law, technology, journalism, and graphic arts.
- Thanks to a partnership between the University of California at San Francisco, the National Academy of Science and the National Science Foundation, the District's science education curriculum has become a National model.
- The U.S. Department of Education recognized our Professional Development Program as one of the five best in the country.
- Extensive class size reduction in the early grades is paying off in dramatic increases in academic achievement. We have increased instructional time in grades 6 through 9, expanded kindergarten to a full-day program, pioneered language immersion programs in several languages and implemented the most academically rigorous high school graduation requirements in the state.
- The District's annual parent conferences have attracted more than 1,100 parents and family members over the past two years. Parent representatives are active members of the Superintendent's cabinet and Parent/Family Resource Centers serves students and their families in the Mission, Bayview Hunter's Point, Chinatown, and the Excelsior.

At SFUSD, we know that these increases are the result of quality, hard and smart working teachers, school site administration, and significant professional development in literacy, math, and science which has lead to improvements in the quality of teaching and learning in the district. In this high stakes accountability arena multiple assessment indicators, both quantitatively and qualitatively, must receive substantial consideration. In the San Francisco story, we believe there are many other pieces of evidence to support the claim of improvement. For example:

- The district annual drop out rate fell from 3.2% in 1992-93 to 2.4% in 1996-97.

- The cohort graduation rate, measured by following how many students starting in eighth grade complete twelfth grade, increased during that same period from 85.3% to 89.3%.
- The number of SFUSD students attending the University of California at Berkeley has increased by nearly 95 percent since 1990.
- Over 75% of the high school graduates attend a college or university.
- Since 1992-93, suspensions and expulsions have dropped as well. The percent of students suspended as a percent of enrollment dropped from 3.63% in 1992-1993 to 2.96% in 1996-97.

Question 1: But how do we know these positive results are indicators of improvements in teaching and learning?

In short, it is because SFUSD has a strong ***accountability system***, which began development in 1983 as a result of a Consent Decree, monitored by the Federal Court. The primary objective of the system is to support data-driven decision making by school administrators and teachers to improve teaching and learning throughout the district. We developed this system because we believe that by analyzing student achievement data and collaboratively reflecting on data, school administrators and teachers can assist schools to meet the “learning imperative” and become successful schools that reach all children (Beck & Murphy, 1996). San Francisco’s accountability system consists of:

- a process of analyzing current conditions at the school; that is, gaining a better understanding of how students are performing now and what their specific learning needs are
- developing a School Site Plan for meeting those learning needs
- implementing those plans
- and monitoring their progress throughout the year.

At the end of the year, every school participates in the district-wide process of evaluating progress on the site-specific objectives and on meeting the district’s goals and the Superintendent’s Priorities.

The data that school decision-makers analyze is not just raw test score data. As part of the accountability system and regular reports generated by the district, test scores are routinely ***disaggregated*** by ethnicity, LEP status, lunch status, Title I, and special education status. Disaggregated data informs decision-makers and helps to identify both areas of strength and gaps in performance for specific groups of students. This data also helps to understand important changes in district demographics and test taking patterns; patterns that can affect test scores. For example, data analysis by the district can answer questions about who is exempted from testing and how many economically disadvantaged students there are in the district.

Question 2: So are specific groups of students not included in the data analysis?

The operating premise is that all children participate in the assessment system. Every student is tested in SFUSD with the exception of kindergarten students and twelfth

graders. In addition, NEP/LEP students who have had less than 30 months of instruction are exempted, and we do not test special education students whose IEP exempts them from taking the CTBS (See Table 1). *The vast majority of those students who are eligible are tested, and the percent tested hasn't changed in the five years with the greatest growth in scores.* (Please see Attachment A for a detailed breakdown of who gets tested.)

Table 1: Percentage of Students Tested

Year	Enrollment	Number Exempt From Testing	Number Eligible For Testing	Students Tested	Percent Eligible Tested	Percent of Enrolled Tested
1992-93	63,374	11,631	51,743	47,183	91.2%	74%
1993-94	62,983	11,136	51,847	46,723	90.1%	74%
1994-95	62,704	15,222	47,482	43,502	91.6%	69%
1995-96	62,918	14,978	47,940	43,708	91.2%	69%
1996-97	63,019	16,029	46,990	42,399	90.2%	67%
1997-98	63,381	15,655	47,326	42,911	90.7%	68%

San Francisco reports both unmatched and matched CTBS scores in its annual Academic Achievement reports. The *unmatched scores* are the performance scores for all the students who took the test in any one test administration. The *matched scores* are scores of those students who were enrolled in the District the year the test was administered and in the previous year. Analysis of the data in this manner gives a clear perspective of the progress of students who have received a full year of instruction in the San Francisco Unified School District.

San Francisco provides intensive analysis of *matched scores*, that is, consecutive years' scores for individual students. Each year, new groups of students are tested, and some groups are no longer tested. For example, a new group of first graders is tested, and a new group of twelfth graders is no longer tested. Mean scores, then, for individual grades, may vary from year to year. For this reason, the district reports results from matched scores, students for whom there are two consecutive years of test results. These results are better estimates of student growth, since they measure each student's learning against their own previous year's score.

The use of matched scores does mean that the number of students for whom the district reports data is smaller than the number of students tested. Not only are first grade students excluded because they have no previous year's score, but so too are students who have taken a "modified" version of the CTBS or took the CTBS in a special setting as indicated on their Special Education IEP.

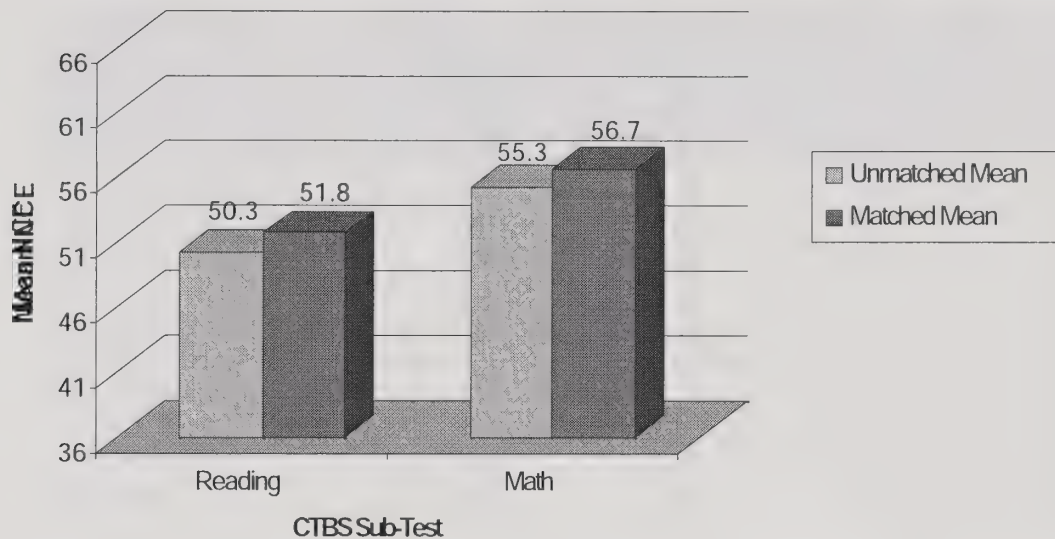
Table 2 shows that of all possible students who could have their scores matched, the SFUSD was able to match approximately ***94% in 1998 -- a percentage that has not changed dramatically in the last six years.*** (Please see Attachment B for a detailed breakdown of who gets tested.)

Table 2: Percentage of Students with Matched Scores

Year	Students Tested	Students with Norm-Referenced Scores	Eligible for Matched Scores	Two Years of Norm-Referenced Scores (Matched)	Percent Matched
1992-93	47,183	45,811	37,722	35,358	93.7%
1993-94	46,723	45,392	37,699	35,253	93.5%
1994-95	43,502	41,664	36,036	33,869	94.0%
1995-96	43,708	41,688	34,282	31,950	93.2%
1996-97	42,399	40,268	33,896	31,824	93.9%
1997-98	42,911	40,498	33,163	31,204	94.1%

In 1997-98, matched scores differed little from scores of all students who took the CTBS Reading Comprehension and CTBS Math Applications and Concepts. The mean performance for all students who took the CTBS in 1997-98 was 50.3 in Reading and 55.3 in Math, while the mean performance for matched score students was 51.8 in Reading and 56.7 in Math (See Chart 2, also See Attachment C for a more detailed breakdown).

**Chart 2: CTBS Mean NCEs for 1997-98:
Unmatched vs. Matched Scores**



Key:

Ranges of Performance	NCEs	National Percentiles
Very low	1-24	1-11
Low	25-35	12-25
Low average	36-44	26-39
Average	45-55	40-60
High average	56-64	61-74
High	65-75	75-88
Very High	76-99	89-99

Question 3: But has the ethnic breakdown of students tested changed? Wouldn't that affect the growth of scores?

The breakdown of students taking the test by ethnicity is roughly equivalent to the ethnic breakdown of the district's enrollment, with only slight over-representation of Chinese students (by 3%) and under-representation of Latino students (by 4%) in the tested group. Moreover, the breakdown of students taking the test by ethnicity has changed very little since 1992-93.

The percentage of students taking the test in 1993 who were African American and Latino is similar to that in 1998 (See Table 3).

Table 3: Ethnic Breakdown of Students Tested

Ethnicity	Percent of District Enrollment in 1993	Percent of Total Tested 1993	Percent of District Enrollment in 1998	Percent of Total Tested 1998
Native American	1%	1%	1%	1%
African American	18%	17%	17%	18%
Chinese	25%	28%	27%	27%
Filipino	8%	8%	7%	7%
Japanese	1%	1%	1%	1%
Korean	1%	1%	1%	1%
Other Non-White	12%	14%	12%	13%
Other White	15%	13%	13%	13%
Latino	20%	18%	21%	19%

Furthermore, the mean performance on the CTBS reading and math increased from 1993 to 1997 *across all ethnic groups* (See Table 4).

Table 4: Performance of Students Tested by Ethnicity

Ethnicity	Reading 1993	Reading 1998	Math 1993	Math 1998
Native American	47.9	47.8	46.9	50.1
African American	38.8	41.0	35.4	39.9
Chinese	51.5	56.2	61	67.8
Filipino	48	50.6	48.2	53.7
Japanese	60.9	63.3	67.6	72.5
Korean	57.9	63.7	63.8	70.3
Other Non-White	46.9	52.1	51.1	58.5
Other White	59.4	63.1	58.9	65.5
Latino	39.3	44.5	37.8	44.6

Question 4: Why exclude Limited English Proficient (LEP) students from taking the test: Are your test scores inflated because you don't test LEP students?

In fact, the total number of LEP students in the school district has been increasing in the past five years, especially in grades one and two. To date, the district has followed state guidelines as well as the National Assessment of Educational Progress guidelines that require testing for Non English Proficient/Limited English Proficient students after 30 months of instruction. This practice is not exclusionary. LEP students with less than 30 months of English instruction may take the test based on the recommendation of their teachers or at a parent's request. Overall, evidence provided to the district from Dr. David Ramirez of the Center for Language Minority Education and Research at California State University-Long Beach suggests that waiting to test students in English until after they've mastered some English is better practice.

There have been changes in the numbers of LEP students tested in the past five years. First, before 1995, the district tested LEP students with more than ten months of instruction. Beginning in 1995, the district adopted the state's guidelines (testing those students with more than 30 months' instruction). The number of students exempted from the test increased from 1,242 to 5,030.

Exempting LEP students from taking the CTBS in all likelihood means that San Francisco's test score growth is *less than it would be if LEP students with 10-30 months instruction had been included in testing because LEP students would tend to show the greatest gains*. In the past three years, overall LEP students tend to show higher gains than other students in the district in both reading and math, as Charts 3 and 4 below and on the following page show.

Chart 4: Math Gains for LEP Students with <30 Months' Instruction

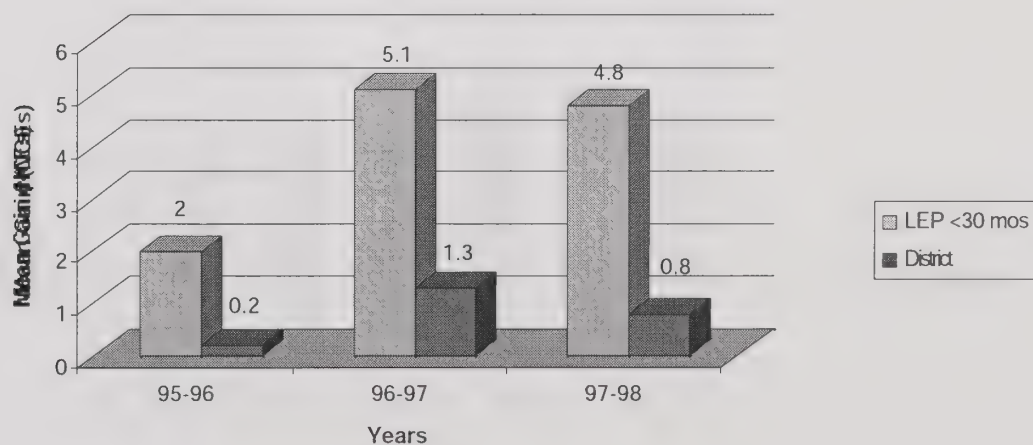
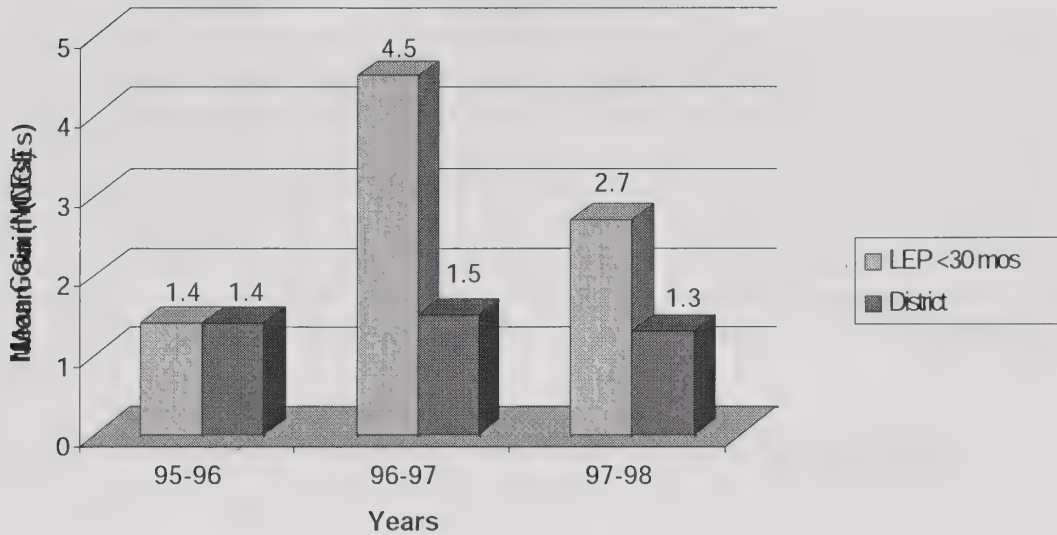
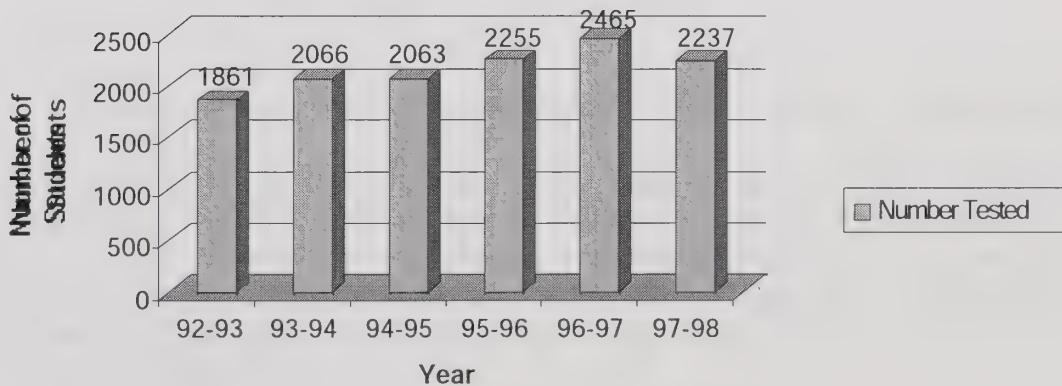


Chart 5: Reading Gains for LEP Students with <30 Months' Instruction



Not all LEP students are exempted from testing altogether. Many students take the SABE, a Spanish-language version of the CTBS. In fact, the numbers of students taking that test has increased by 20% since 1993. This past year, 2,237 students took the SABE (See Chart 6).

Chart 6: Number of Students Taking the SABE

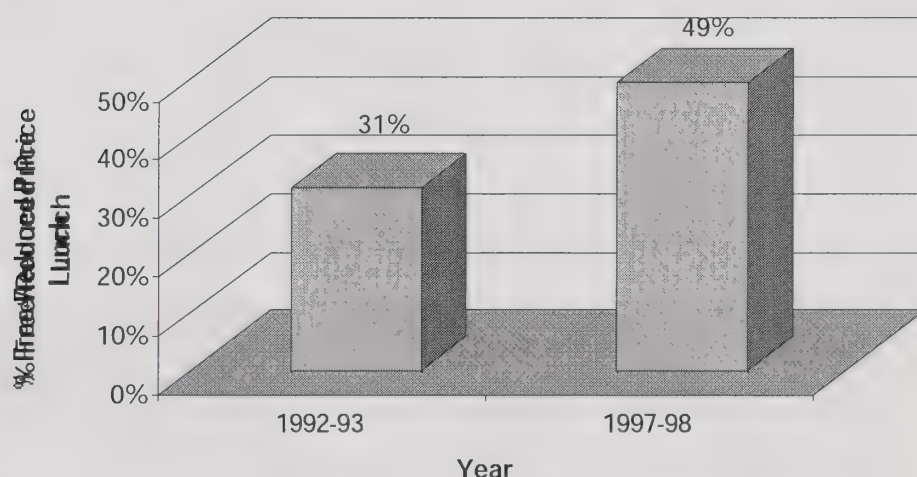


Question 5: Is the socioeconomic status of students in the district increasing, resulting in higher test scores?

Test scores in a district may increase when the socioeconomic status of district's students rises, because there is a well-documented high correlation between test scores and income. Regretfully, students in the SFUSD are becoming more economically disadvantaged as a whole.

The percentage of students receiving free or reduced price lunch, a good indicator of socioeconomic status *increased* dramatically between 1992-93 and 1997-98 from 31% (n=20,022) to 49% (n=29,733) (See Chart 7). These data suggest that, if anything, San Francisco students on the whole are more economically disadvantaged today than they were five years ago. From this data, one would predict lower, not higher, test scores district-wide.

Chart 7: Free & Reduced Lunch Recipients: 1992-93 Compared to 1997-98



Question 6: What external evidence do we have that SFUSD students are improving?

Performance on the CTBS for the last 6 years is not the only achievement assessment test on which SFUSD students have shown improvement. Students' SAT scores have steadily increased by 10 points for Verbal and 8 points for Math from 1993 to 1998 (See Chart 8). In addition, student participation as well as the percentage of students receiving honor rewards on the Golden State Exams have increased from 1995 to 1998 (See Table 5).

Chart 8: District SAT Scores from 1992-93 to 1996-97

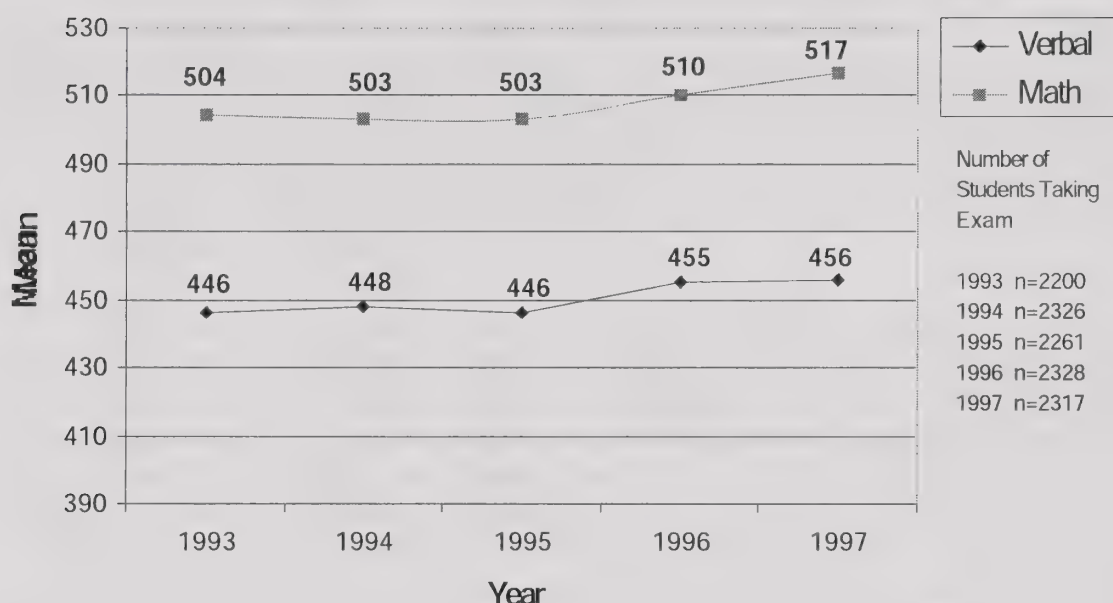


Table 5: District Honor Recipients on the Golden State Exam

Year	No. of Participants	Total No. Awarded	Total Percent Receiving Honor Awarded
1995	5,355	2,171	41%
1996	7,022	2,937	42%
1997	8,090	3,590	44%
1997	10,4160	4,160	41%

Question 6: What plans are there for meeting the challenge of continuous improvement?

Each year the District Goals, Philosophical Tenets, and *Superintendent's Priorities* guide planning for the district. These priorities are set based on data from previous years presented in a report called *The Critical Path*. The Critical Path presents trends in district-wide student achievement and behavioral data, including the performance and gains by school level, ethnicity, grade and quartile on the reading and math sub-tests of the CTBS. Chief among those priorities this year is raising the achievement level of African American and Latino students and English Language Learners to the 50th percentile by 1999. In 1992, a large number of these students were over-represented in the bottom quartile of reading and math. Now that the district has reduced the numbers of students in the bottom quartile, the focus is on raising the achievement level of second and third quartile students. Additional priorities for the Superintendent relate to enhancing school-to-work programs, improving gifted services and achievement, strengthening language acquisition initiatives, creating full-service

“one-stop” schools, improving buildings and maintenance, and improving educational technology.

The African American Community Education Partnership Summit focusing on reading and mathematics achievement was held in November 1997. Follow-up workshops have been held to develop comprehensive strategies for continuous improvement. In April, 1999 the SFUSD sponsored the Latino Education Summit to focus on the academic achievement of Latino students. At these summits, parents and other members of the community had the opportunity to express their concerns and to provide recommendations about the educational opportunities of African American and Latino children in the District.

The District’s continuous improvement effort is carefully planned upon extensive analysis of multiple sources of data including CTBS results, performance based assessments matched to world-class standards, PSAT/SAT results, Golden State exams, advanced placement course work, and observation of classroom practice. The analysis of data continually references the District Standards pre-K through 12. The District Standards, published in multiple languages, define rigorous content and performance benchmarks that enable all students to graduate from SFUSD meeting college preparatory requirements. Curriculum guides and professional development based on rigorous standards help teachers understand how to scaffold learning so that all students’ needs are addressed. The district initiatives are supported by partnerships with institutions of higher education, such as, University of California, Berkeley, Stanford, San Francisco State University, and City College of San Francisco. Also, there are partnerships that support specific areas of the curriculum, such as, the Exploratorium and the California Academy of Sciences in the areas of math and science. The District Professional Development Initiative focuses on student’s academic achievement supporting schools continuous work and reflection on high expectations, rigorous standards, and learning through the examination of student work.

ATTACHMENT A

Districtwide CTBS/4 Test Administrations from 1992-93 through 1997-98

	Enrollment	Categories Exempt	Number Exempt	Number Eligible	Students Tested	% Tested
1992-93	63,374	Total	11,631	51,743	47,183	91.20%
		Kindergarten	5,332			
		Grade 12	4,011			
		Ungraded	657			
		LEP < 10	1,631			
		Special Ed	NA			
1993-94	62,983	Total	11,136	51,847	46,723	90.10%
		Kindergarten	5,147			
		Grade 12	3,996			
		Ungraded	751			
		LEP < 10	1,242			
		Special Ed	NA			
1994-95	62,704	Total	15,222	47,482	43,502	91.60%
		Kindergarten	5,088			
		Grade 12	3,837			
		Ungraded	738			
		LEP < 30	5,030			
		Special Ed	529			
1995-96	62,918	Total	14,978	47,940	43,708	91.20%
		Kindergarten	5,168			
		Grade 12	3,996			
		Ungraded	701			
		LEP < 30	5,113			
		Special Ed	NA			
1996-97	63,019	Total	16,029	46,990	42,399	90.20%
		Kindergarten	5,115			
		Grade 12	3,590			
		Ungraded	521			
		LEP < 30	6,390			
		Special Ed	413			
1997-98	63,381	Total	16,055	47,326	42,911	90.70%
		Kindergarten	5,190			
		Grade 12	3,812			
		Ungraded	601			
		LEP < 30	5,654			
		Special Ed	798			

ATTACHMENT B

Districtwide CTBS/4 Test Administrations from 1992-93 through 1997-98

	Students Tested	Special Cases	No Special Cases	Students with Norm-Referenced Scores (Unmatched)		Ineligible for Match	Eligible for Matched	Two Years of Norm-Referenced Scores (Matched)	% Matched
1992-93	47,183	Total	1,372	45,811	First Tested	8,089	37,722	35,358	93.7%
		Modified	773		Grade 1	4,548			
		Spec Setting	599		LEP	848			
					New to SF	2,693			
1993-94	46,723	Total	1,331	45,392	First Tested	7,693	37,699	35,253	93.5%
		Modified	792		Grade 1	4,615			
		Spec Setting	539		LEP	998			
					New to SF	2,080			
1994-95	43,502	Total	1,838	41,664	First Tested	5,628	36,036	33,869	94.0%
		Modified	1,151		Grade 1	3,207			
		Spec Setting	687		LEP	593			
					New to SF	1,828			
1995-96	43,708	Total	2,020	41,688	First Tested	7,406	34,282	31,950	93.2%
		Modified	1,350		Grade 1	3,316			
		Spec Setting	670		LEP	2,298			
					New to SF	1,792			
1996-97	42,399	Total	2,131	40,268	First Tested	6,372	33,896	31,824	93.9%
		Modified	1,669		Grade 1	2,919			
		Spec Setting	462		LEP	1,809			
					New to SF	1,644			
1997-98	42,911	Total	2,413	40,498	First Tested	7,335	33,163	31,204	94.1%
		Modified	1,928		Grade 1	3,008			
		Spec Setting	485		LEP	2,711			
					New to SF	1,616			

ATTACHMENT C

Districtwide CTBS/4 Test Administrations from 1992-93 through 1997-98

	Unmatched Scores				Matched Scores			
	Number Read	Read NCE	Number Math	Math NCE	Number Read	Read NCE	Number Math	Math NCE
1992-93								
District	45,523	45.0	45,349	48.4	34,708	47.5	34,469	50.0
LEP	12,211	31.2	12,196	38.7	8,187	32.9	8,194	38.5
Title I	14,588	33.4	11,207	30.5	14,588	33.4	11,207	30.5
Special Ed	3,513	30.3	3,482	28.7	2,370	32.5	2,327	30.3
Consent Decree	9,982	39.3	9,957	43.0	7,646	41.4	7,567	43.7
1993-94								
District	45,071	45.1	44,832	48.7	34,864	47.5	34,590	50.3
LEP	12,111	31.4	12,007	38.6	8,133	33.0	8,029	38.5
Title I	13,839	32.5	12,032	31.0	13,839	32.5	12,032	31.0
Special Ed	3,337	30.9	3,316	28.6	2,793	32.1	2,759	29.6
Consent Decree	10,053	40.1	10,050	43.5	7,729	41.9	7,710	44.5
1994-95								
District	41,387	47.7	41,331	51.7	33,542	49.2	33,342	52.8
LEP	8,809	34.5	8,787	41.1	6,997	34.7	6,879	40.4
Title I	12,978	34.1	11,359	33.0	12,978	34.1	11,359	33.0
Special Ed	2,992	33.3	2,964	32.3	2,408	34.3	2,374	33.4
Consent Decree	8,954	43.3	8,958	47.1	7,362	44.1	7,370	47.1
1995-96								
District	41,418	48.0	41,301	51.9	31,646	50.3	31,455	53.7
LEP	9,419	36.5	9,436	43.5	5,752	36.5	5,690	42.2
Title I	11,922	34.5	4,968	40.6	11,922	34.5	10,275	32.7
Special Ed	3,329	35.9	3,284	35.1	2,570	37.9	2,526	37.3
Consent Decree	9,008	42.8	9,006	46.1	6,977	44.6	6,949	47.1
1996-97								
District	39,863	49.3	39,736	53.9	31,436	50.8	31,296	55.3
LEP	9,271	38.6	9,215	45.6	6,482	38.8	6,443	45.6
Title I	13,447	37.2	13,376	39.2	12,600	37.1	11,359	37.6
Special Ed	2,545	35.0	2,512	35.5	1,888	37.0	1,855	37.7
Consent Decree	9,112	43.4	9,044	47.2	7,155	44.5	7,109	47.8
1997-98								
District	40,279	50.3	40,167	55.3	30,939	51.8	30,818	56.7
LEP	8,802	38.4	8,765	45.6	4,983	36.5	4,941	42.9
Title I	13,014	38.4	12,964	41.2	10,424	36.7	9,388	37.1
Special Ed	22,293	37.3	2,266	38.0	1,638	39.4	1,611	40.5
Consent Decree	9,454	44.3	9,436	48.2	7,291	45.4	7,258	49.0

ATTACHMENT C1

Districtwide SABE Test Administrations from 1992-93 through 1996-97

	Unmatched Scores				Matched Scores			
	Number Read	Read NCE	Number Math	Math NCE	Number Read	Read NCE	Number Math	Math NCE
1992-93	1,845	29.1	1,856	28.9	1,095	27.2	1,096	28.8
1993-94	2,052	30	2,052	30.5	1,096	29.9	1,110	31.2
1994-95	2,063	30.1	2,063	29.7	1,291	29.7	1,301	31.6
1995-96	2,245	33.7	2,255	33.9	1,111	31.2	1,115	34.1
1996-97	2,452	36.1	2,457	36.1	1,514	32.9	1,511	36.6
1997-98	2,218	38.6	2,206	39.4	1,465	35.5	1,454	39.3

Why We Do Not Use English-Language Achievement Tests To Assess English Language Learners

From the *NCME Code of Professional Responsibilities in Educational Measurement* (National Council on Measurement in Education):

Section 4. Responsibilities of Those Who Administer Assessments

4.7. avoid any conditions in the conduct of the assessment that might invalidate the results.

4.11. avoid actions or conditions that would permit or encourage individuals or groups to receive scores that misrepresent their actual levels of attainment

Section 7. Responsibilities of Those Who Educate Others About Assessment

7.5 avoid administering any assessment that is not part of the evaluation of student performance in a course if the administration of the assessment is likely to harm any student

7.6 Avoid using or reporting the results of any assessment that is not part of the evaluation of student performance in a course if the use or reporting of results is likely to harm any student

The establishment of such clear standards for testing and assessment aimed at preventing misrepresentation of actual achievement and preventing harm to those being tested has been hard-won in the United States. The history of assessment of intelligence in this country has a troubling past. At the turn of the century, reports Gould (1981), new immigrants to Ellis Island were inspected as to their intelligence and returned to their country of origin if they were deemed “feeble-minded.” One inspector tested thirty-five Jews, twenty-two Hungarians, fifty Italians, and forty-five Russians. These groups could not be regarded as random samples, since government officials had already sent some immigrants home as “defective.” Binet tests showed that 83% of Jews, 80% of Hungarians, 79% of Italians, and 87% of Russians were feeble-minded, that is, below age twelve on the Binet scale of intelligence. The inspector himself was flabbergasted: could anyone be made to believe that four-fifths of any nation were morons?

In recent years, court cases have ensured that many of the injustices caused by the mismeasurement of intelligence and achievement do not recur. For instance, *Lau vs. Nichols* (1974) required states to provide equal opportunity for language minority students in public schools, including, if necessary, academic instruction in students’ native language. In a subsequent memorandum, *Diana vs. State Board of Education* stated that districts must meet the linguistic and cultural needs of students in the assessment process and test in the child’s native language so that no child will be placed into special education only on the basis of limited ability to speak English.

In San Francisco, there are four primary reasons why we do not test all English Language Learners (ELL) on English achievement tests. The reasons are as follows:

1. Invalid for diagnosis of student achievement
2. Causes harm to students
3. Wastes public money
4. Violates established testing standards

A detailed account for each reason is presented below. Each reason includes codes/standards from testing practices handbooks, manuals, and citations from research in the field. Lastly, research findings about best practices in teaching and assessing language minority students are presented.

Reason 1: Testing students who do not understand English does not provide a valid baseline or diagnostic tool for interpreting student achievement.

- Failing to assess content knowledge in students' native language is "likely to underestimate students' academic achievement" (Anstrom, 1997). Underestimation occurs because *standardized tests for students who are non-native speakers of English inevitably not only are tests of the domain being tested but are also tests of English proficiency* (American Psychological Association, 1985, p. 73; Figueroa, 1987; Anstrom, 1997). Achievement tests have been found to be "hypersensitive to language background" (Valdes & Figueroa, 1994). Giving a mathematics test to a non-native speaker of English is unlikely, then, to give a valid, meaningful indicator of how well they are doing in mathematics.
- Identical low test scores do not mean the same thing when one student knows the language and another student does not. It is not the numbers themselves, but what the numbers represent that is of interest to teachers and researchers. When the same number represents two different phenomena, the results are of little value (Grimm & Yarnold, 1995).
- Test scores at the bottom of a scale have little diagnostic value (CTB/McGraw-Hill, 1990, p.88). When students score close to the bottom, it is not known whether the student does not understand English or does not know the content being assessed. Imagine giving an adult who speaks, reads and writes only English a test of reading comprehension in Russian. If the adult got a low score, one could not necessarily conclude that the adult did not understand how to read, only that the adult cannot read Russian. Knowing that he got a low score would not provide information to either the adult or a teacher on how to improve the adult's reading comprehension, unless the test identified that the problem is that the adult does not know Russian. The SAT-9 and other tests like it do not provide such information.

Reason 2: Testing students in English who have no knowledge of English is likely to cause harm to students, particularly if they are provided feedback that indicates they perform poorly on an important measure of academic achievement.

- Researchers have found there are strong links between the provision of feedback on academic achievement, academic self-confidence, and future academic performance (Leach & Ingram, 1989; Schunk & Swartz, 1993). Negative feedback on task performance has been shown to have effects on students liking of similar tasks, even when they are easy (Tang & Sarsfield, 1989). If accurate, negative feedback could help teachers and students identify what skills they need to develop. Providing students with extremely negative and inaccurate feedback—that is, feedback that is likely to underestimate their achievement—may result in less interest in academic related tasks and lower academic performance in the future.
- Members of a stigmatized social group may protect their overall self-esteem by devaluing those domains in which they are lower performing (Crocker & Major, 1989). When students see ways that members of their ethnic group are systematically denied to access to educational opportunities and other social goods, they perform lower than their peers (Ogbu, 1978; 1992). LEP students historically have often been isolated from other students in school districts (Donato *et al.*, 1993). If students receive extremely low scores, they may decide that doing well in school is not a goal they will seek to attain.

Reason 3: Given that testing NEP/LEP students will yield invalid results, testing wastes valuable public money that could be spent on improving instruction.

- Last year, 6,390 LEP students in San Francisco were exempt from testing, in conformity with the state's criteria. The additional cost to California taxpayers of the STAR program just for testing these students in San Francisco would be at least \$42,493.50. This figure fails to take into consideration overhead costs of administering the tests in schools, costs absorbed by the school district's own budget. Statewide, some 1.3 million students are classified as LEP. If all are tested, the cost would be 8.6 million dollars, or 25% of the state's budget for the STAR program.

Reason 4: Testing all students in English, regardless of their proficiency in English, violates equal opportunity for academic achievement as well as known standards for administration and use of educational testing.

The three parts of any instructional program are assessment, instruction, and evaluation. Testing provides the diagnostic information for planning instructional programs and for evaluation of student achievement and success in the program. Therefore, testing Limited English Proficient Students in English before they have acquired the comparable English skills is not educationally sound, and it does not comply with Federal and State law that require equal opportunity for academic achievement.

- From the *NCME Code of Professional Responsibilities in Educational Measurement* (National Council on Measurement in Education):

Section 4. Responsibilities of Those Who Administer Assessments

- 4.7. avoid any conditions in the conduct of the assessment that might invalidate the results.
- 4.11. avoid actions or conditions that would permit or encourage individuals or groups to receive scores that misrepresent their actual levels of attainment

Section 7. Responsibilities of Those Who Educate Others About Assessment

- 7.5 avoid administering any assessment that is not part of the evaluation of student performance in a course if the administration of the assessment is likely to harm any student
- 7.6 Avoid using or reporting the results of any assessment that is not part of the evaluation of student performance in a course if the use or reporting of results is likely to harm any student

Implications: Assessing NEP and LEP students in English is likely to invalidate the results of the test, as explained in the first section: tests in specific content areas, such as mathematics, are unlikely to be measuring mathematics achievement alone. Rather, they are measuring, at least in part, English proficiency, which is not the intent of a mathematics achievement test. Thus, students' scores on such tests are likely to misrepresent their actual level of achievement. Similarly, administering tests to students in a language they do not know at all is likely to cause harm to students, as is the reporting of results, especially when the school, district, and state climate all point to the importance of doing well on such tests as indicators of student achievement.

- From the *Code of Fair Testing Practices in Education* (comprised of: the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education)

Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test taking population.

Section B. Interpreting Scores

- 10. Test users should interpret scores taking into account any major differences between the norms or comparison groups and the actual test takers. Also take into account any differences in test administration practices or familiarity with the specific questions in the test.

Section C: Striving for Fairness

- 15. Review the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Evaluate the extent to which performance differences may have been caused by inappropriate characteristics of the test.

Implications: The SAT-9 reports that the only LEP students in their norm sample were those who would normally be tested as part of their educational program; therefore, the validity and reliability of using such tests for all LEP students is unknown (SAT-9 Technical Data Report, 1996). It is likely as well that performance differences between LEP and non-LEP students are largely due to “inappropriate characteristics of the test,” namely that the test was given in a language unfamiliar to the student.

- From Standards for Testing Language Minority Students, in *Standards for Educational and Psychological Testing* (the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education)

Standard 13.1 For non-native English speakers or for speakers of some dialects of English, testing should be designed to minimize threats to test reliability and validity that may arise from language differences.

Standard 13.2 Linguistic modifications recommended by test publishers should be described in detail in the test manual.

Standard 13.7 English language proficiency should not be determined solely with tests that demand only a single linguistic skill.

Implications: Again, the validity and reliability of test scores are seriously compromised when students cannot read the language of the test at all. Even as a measure of English proficiency, one test is insufficient to assess student achievement (Garcia, 1991; Genessee & Upshur, 1996). As the Standards state, “It is important to recognize the limits of interpretations drawn from tests developed without due consideration of the influence of the linguistic characteristics of some test takers” (APA, 1985, p. 73).

- From the *National Assessment of Educational Progress*, inclusion rules for taking the NAEP tests:

Students are included if:

Student has received academic instruction primarily in English for at least three years,
OR

Student has received academic instruction in English for less than three years, if school staff determine that the student is capable of participating in the assessment in English,
OR

Student, whose native language is Spanish, has received academic instruction in English for less than three years, if school staff determine that the student is capable of participating in the assessment in Spanish (if available)

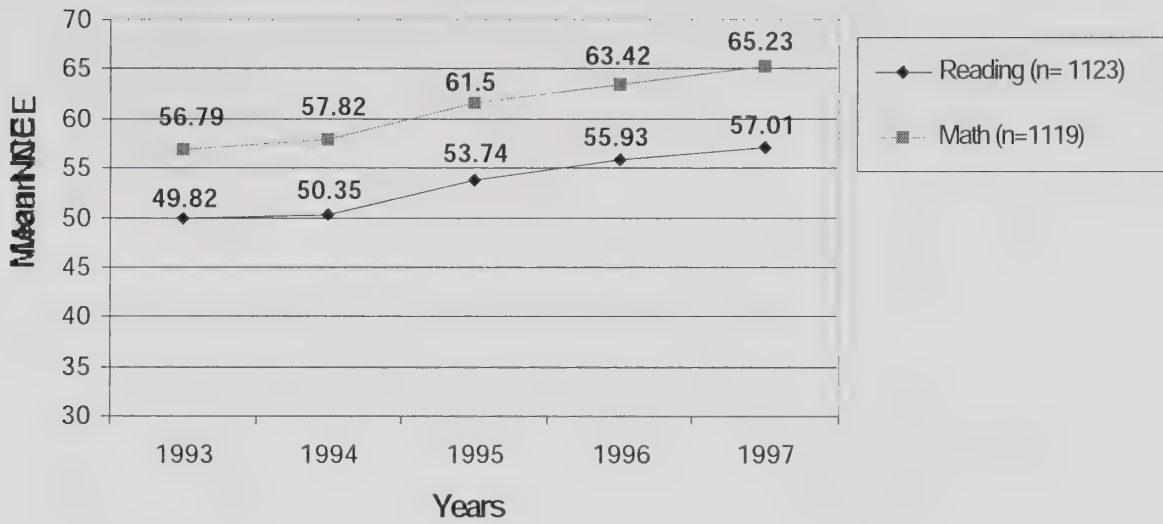
Implications: NAEP sets the standard of being able to understand English as the basis for inclusion in testing. The NAEP standards are consistent with the state of California’s own requirements as of last year: students with more than 30 months instruction are included.

Research Findings from Studies of English Language Learners

Bilingual research findings suggest that enrolling students in two-way or late-exit bilingual programs is more successful than English-only instruction, and by implication, English-only testing upon immediate enrollment. Other assessment tools are more appropriate with NEP/LEP students with less than 30 months instruction. The following research supports the above claim:

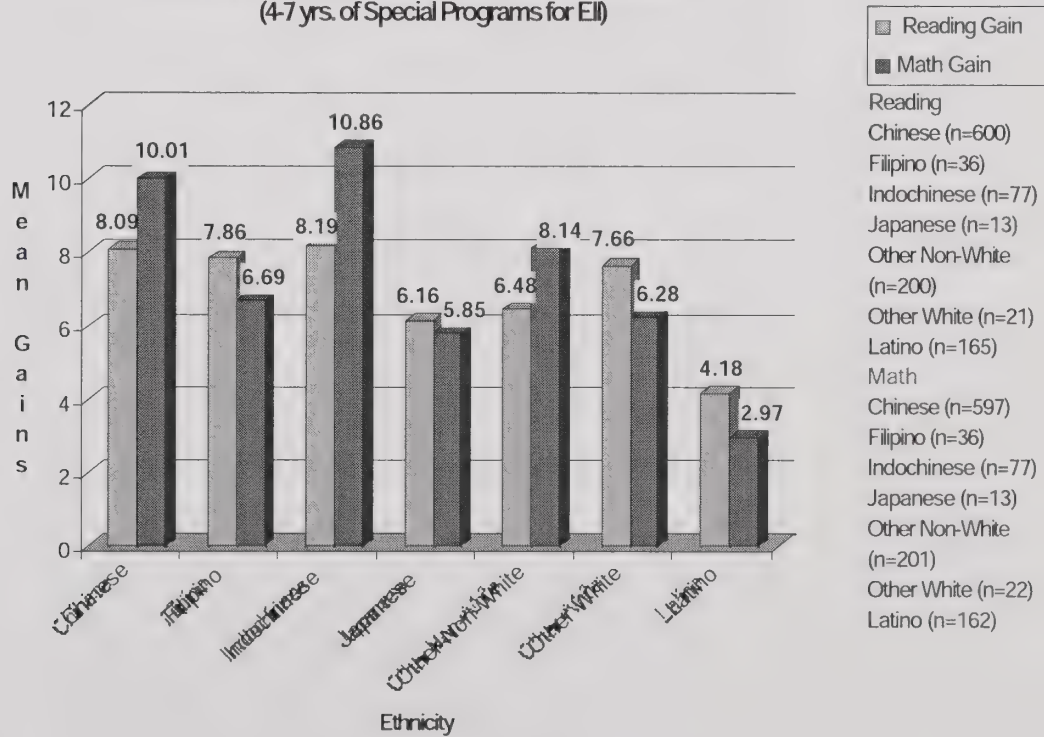
- Research on bilingual education has shown that teaching primarily in students' native language (sometimes called L1) in early grades does not slow down the acquisition of English-language skills or literacy (Crawford, 1997; Hakuta & Diaz, 1985). In fact, students in "late-exit" bilingual programs--where English is introduced in kindergarten (10%) and gradually increased by fourth grade before exiting the program after grade 6—**approach national norms by 6th grade** (Ramirez, Yuen, & Ramey, 1991).
- In a related study, Thomas and Collier (1996) found that English-language learners that had support and instruction in their native language took between 4-7 years on average to reach the 50th percentile in English reading, while students with no support in their native language took between 7-10 years on average to reach the 50th percentile.
- Thomas and Collier (1996) also found that students who received early-exit bilingual programs (where academic instruction takes place for half a day in each language and transition to English-only classrooms occurs within 2-3 years) and ESL instruction (no instruction in native language) were much **less likely to ever reach the 50th percentile in English reading** than students who were in late-exit bilingual programs or two-way bilingual programs.
- An analysis of SFUSD's own scores shows that on average, **redesignated Fluent English Proficient (FEP) students who have had between 4-7 years of native language instruction are well above national norms in reading and math and are rising each year.** These students were selected for analysis because they are students who would be considered "late-exit" by Ramirez, Yuen, & Ramey, 1991), and would be predicted to show the greatest benefit of native language instruction on English achievement after redesignation as FEP.

CTBS Matched Scores For Late Exit FEP Students
(4-7 yrs. of Special Programs for ELL)
Redesignated on or before 1993



- An analysis of SFUSD data further shows that this **upward trend holds true for every ethnic group.**

CTBS Gains For 1993 to 1997 For Late Exit FEP Students
(4-7 yrs. of Special Programs for ELL)



Other means of assessment other than English language achievement tests provide more valid means of remaining accountable for the learning of English Language Learners. Assessment of English Language Learner achievement across a variety of domains- not just English proficiency- requires taking into account both the child's languages. Researchers have identified that a portfolio approach, in which student work in both languages is collected and analyzed over time, is the most valid approach to assessing English Language Learners progress in school (Anstrom, 1997; Navarette, Wilde, Nelson, Martinez, & Hargett, 1990; Valdez-Pierce & O'Malley, 1992).

Research Cited

- American Psychological Association. (1985). Standards for educational and psychological testing. Washington, DC: APA.
- Anstrom, K. (1997). Academic achievement for secondary language minority students: Standards, measures, and promising practices. Washington, DC: National Clearinghouse for Bilingual Education.
- CTB/McGraw-Hill. (1990). Comprehensive Test of Basic Skills, fourth edition: Test coordinator's handbook. Monterey, CA: MacMillan/McGraw-Hill.
- Crawford, J. (1997). Best evidence: Research foundations of the Bilingual Education Act. Washington, DC: National Clearinghouse for Bilingual Education.
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. Psychological Review, 96, 608-630.
- Donato, R. *et al.* (1993). Segregation, desegregation, and integration of Chicano students: Problems and prospects. In R.R. Valencia (Ed.), Chicano school failure and success: Research and policy agendas for the 1990s. New York: Falmer Press.
- Figueroa, R. A. (1987). Special education assessment of Hispanic pupils in California: Looking ahead to the 1990s. Sacramento: California Department of Education, Office of Special Education.
- Garcia, G.E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. Research Reading Quarterly, 26, 371-392.
- Genessee, F., & Upshur, J.A. (1996). Classroom-based evaluation in second language acquisition. New York: Cambridge University Press.
- Gould, S.J. (1981). The mismeasure of man. New York: Norton.
- Grimm, L.G., & Yarnold, P.R. (1995). Reading and understanding multivariate statistics. Washington, DC: APA.
- Hakuta, K., & Diaz, R.M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K.E. Nelson (Ed.), Children's language (Vol. 5). Hillsdale, NJ: Erlbaum.
- Leach, D.J., & Ingram, K.L. (1989). The effects of information and feedback on teachers' classroom behavior and students' academic engaged time. Educational Psychology, 9, 167-184.
- Navarrette, C., & Guskee, C. (1996). A guide to performance assessment for linguistically diverse students. Albuquerque, NM: Evaluation Assistance Center-West.
- Navarrete, C., Wilde, J., Nelson, C., Martinez, R., & Hargett, G. (1990). Informal assessment in educational evaluation: Implications for bilingual education programs. FOCUS Occasional Papers in Bilingual Education from the National Clearinghouse for Bilingual Education.
- Ogbu, J. (1978). Minority education and caste. New York: Academic Press.
- Ogbu, J. (1992). Understanding cultural diversity and learning. Educational Researcher, 21, 5-14, 24.
- Ramirez, J.D., Yuen, S.D., & Ramey, D.R. (1991). Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for bilingual-minority children. San Mateo, CA: Aguirre International.
- Stanford Achievement Test Series, Ninth Edition: Technical data report. (1996). San Antonio: Harcourt Brace.
- Schunk, D.H., & Swartz, C.W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. Contemporary Educational Psychology, 18, 337-354.

- Tang, T.L-P., & Sarsfield, L.T. (1989). The effects of self-esteem, task-label, and performance feedback on goal setting, efficacy, and task performance. ERIC Document 312 525.
- Thomas, W. P. & Collier, V.P. (1996). Language minority student achievement and program effectiveness. Fairfax, Vs.: Center for Bilingual/Multicultural/ESL Education, George Mason University.
- Valdés, G., & Figueroa, R.A. (1994). Bilingualism and testing: A special case of bias. Norwood, NJ: Ablex.
- Valdez-Pierce, L. & O'Malley, J. M. (1992). Performance and portfolio assessment for language minority students. FOCUS Occasional Papers in Bilingual Education from the National Clearinghouse for Bilingual Education.

U.C. BERKELEY LIBRARIES



C124920463